Intelligent Thermal Control Strategy Based on Reinforcement Learning for Space Telescope

Yan Xiong,[*] Liang Guo,[†] Yong Huang,[‡] and Liheng Chen[§]
Chinese Academy of Sciences, 130033 Changchun, People's Republic of China

DOI: 10.2514/1.T5774

In this study, a thermal model of a space telescope is established in Simulink. An intelligent autonomous thermal control strategy based on actor-critic reinforcement learning (RL) for proportional–integral–derivative (PID) parameter adaptive self-tuning, called RL PID, is proposed. This control strategy enables the PID thermal controller to adaptively tune the PID parameters to achieve stable and precise temperature control. A single radial basis function (RBF) neural network is applied to simultaneously approximate the strategy function of the actor and the value function of the critic. The actor maps the system state to PID parameters, and the critic evaluates the output of the actor and generates a temporal difference (TD) error. Based on the architecture of the actor-critic RL algorithm and the TD error performance index, a design flow chart of RL PID is made. Both theoretical and experimental results show that RL PID can achieve a temperature control precision of 0.01°C, and that the steady-state error is reduced by 50 and 75% in the simulation and 50 and 67% in the experiment compared with those of the traditional PID controller and the traditional switch controller, respectively. RL PID has better reliability, more robustness, and a faster response.

Nomenclature

a_{MLI}	=	absorption coefficient of multilayer
c_i	=	specific heat capacity of node i
D_{ij}	=	heat transfer coefficient from node i to node j
E_{ij}	=	radiative heat transfer coefficient from node i to node j
k	=	sampling number
m_i	=	quality of node i
P_{atm}	=	atmospheric pressure
P_{Heat}	=	heating power
PID	=	proportional-integral-derivative
Q_i	=	internal and external heat sources of node i
RBF	=	radial basis function
RL	=	reinforcement learning
RLPID	=	an intelligent autonomous thermal control strategy based on actor-critic reinforcement learning
SAM	=	stochastic action modifier
T	=	sampling period
T_i	=	thermal dynamic temperature value of node i
T_{inside}	=	temperature inside the incubator
T_{outside}	=	temperature outside the incubator
V_{Sup}	=	volume
α	=	weighted coefficients of immediate reward
α_A	=	learning rate of actor
α_C	=	learning rate of critic
β	=	weighted coefficients of immediate reward
δ_{TD}	=	temporal difference
η_μ	=	learning rate of center
η_σ	=	learning rate of width
γ	=	discount factor
ε	=	tolerant error band
ε_{MLI}	=	emission coefficient of multilayer
$\varepsilon_{\mathrm{Sup}}$	=	emission coefficient

I. Introduction

TEMPERATURE is the main factor affecting the performance of space telescopes. The spatial temperature uniformity and keeping temperature change rate are critical to the precision thermal control of space telescope. Proper thermal control strategies are crucial to controlling the temperature of some key components within their proper ranges to guarantee long-term stable operation. The requirements for thermal control have increased with increasing resolution of space telescopes [1,2]. Although the temperature control precision of some space telescopes can reach one-thousandth of a degree Celsius, much higher precision is needed.

High-precision temperature control can be achieved through a variety of methods. Proportional–integral–derivative (PID) thermal control is widely used for space telescopes because of its suitable static and dynamic characteristics, simple implementation, and high robustness.

PID thermal control was applied to meet ±0.05°C temperature precision for the optical assembly of Envisat [3]. An adaptive PID thermal control algorithm was applied to the Swift satellite thermal controller, which is mainly used for active thermal control of an X-ray telescope, with control precision of higher than 0.1°C achieved [4–6]. The US Jet Propulsion Laboratory developed a digital temperature control system for space that uses a platinum resistance thermometer as a temperature sensor and a PID temperature control algorithm to adjust the temperature control power [7]. The system provides fault-tolerant, multizone temperature control with temperature control precision of 0.1°C. The far-infrared optical system of the Herschel and Planck satellites [8] and the thermal control loop of the star sensor adopt a proportional–integral (PI) thermal control algorithm. The CALIPSO satellite thermal control heater [9] is controlled by a programmable logic controller, and also uses a PI control algorithm. The key parameters of a PID controller influence temperature control precision in the traditional control process. And the traditional PID thermal controller needs to manually tuning the parameters, which will take a lot of time and it cannot be adaptively adjusted online. Researchers have thus proposed a number of PID parameter tuning methods, such as fuzzy adaptive PID control [10,11], neural network adaptive PID control [12], and evolutionary algorithm adaptive PID control [13]. Fuzzy adaptive PID control requires a large amount of prior knowledge. Because neural network adaptive PID generally

Received 25 March 2019; revision received 17 May 2019; accepted for publication 15 July 2019; published online 20 August 2019.

[*] Ph.D. Candidate, Thermal Control Group, Changchun Institute of Optics, Fine Mechanics and Physics; also University of Chinese Academy of Sciences, 100049 Beijing, People's Republic of China; xiongyan16@mails.ucas.ac.cn.

[†] Associate Professor, Thermal Control Group, Changchun Institute of Optics, Fine Mechanics and Physics; guoliang@ciomp.ac.cn (Corresponding Author).

[‡] Associate Professor, Thermal Control Group, Changchun Institute of Optics, Fine Mechanics and Physics; huangyong@ciomp.ac.cn.

[§] Professor, Thermal Control Group, Changchun Institute of Optics, Fine Mechanics and Physics; chenliheng3@163.com.

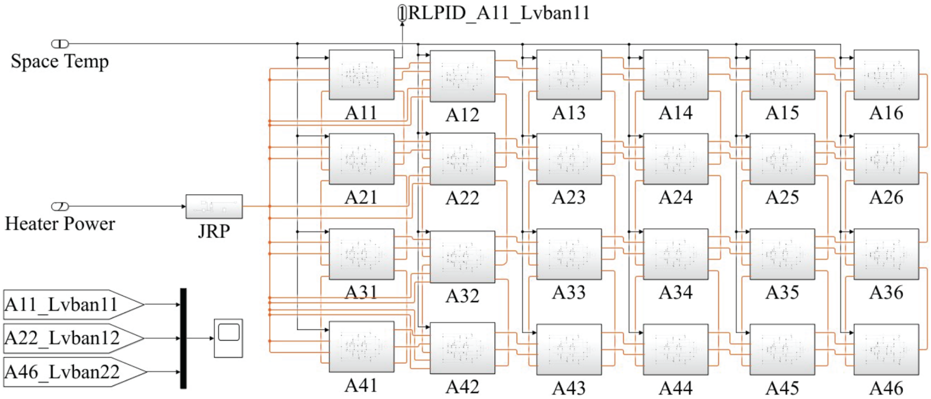


Fig. 1 Thermal model of space telescopes in Simulink.

uses supervised learning to optimize parameters and its teacher signal is difficult to obtain, its practicability is relatively poor. Less prior knowledge is required for evolutionary algorithm adaptive PID control; however, it has a long calculation time and it is difficult to apply it for real-time control [14–16].

To realize high-precision temperature control for space telescopes, an intelligent autonomous thermal control strategy based on reinforcement learning (RL) for PID parameter adaptive self-tuning, called RL PID, is proposed in the present study. RL PID is different from a traditional supervised learning algorithm in that supervised learning adopts a reward and punishment method, whereas RL uses an agent to obtain empirical knowledge according to interactions with the environment. The RL agent first actively explores the environment and then evaluates the results of the exploration based on the optimized controller. This enables unsupervised online learning without a model. The actor-critic learning algorithms proposed by Barto et al. is one of the most important RL methods [17,18]. It is a systematic method for simultaneously finding the optimal action and the expected value in real time. Actor-critic learning algorithms have been widely used for various tasks related to artificial intelligence, robot planning and control, and optimization and scheduling. In the present study, an actor-critic learning radial basis function (RBF)-network-based adaptive PID intelligent thermal controller for space telescopes is proposed. The actor-critic learning algorithms adaptively tune the PID parameters. The controller is online adapted according to system dynamics.

The remainder of this paper is organized as follows. Section II presents a thermal model of a thin plate unit in Simulink. Section III shows the process of RL PID. Sections IV and V, respectively, describe simulation and experimental results. Finally, Sec. VI shows the conclusions of this study.

II. Thermal Model of Space Telescopes

Before RL PID can be applied, it is necessary to first analyze the thermal model of space telescopes. The thermal physics model of space telescopes is very complex. The node network method [19–22] is mainly used for modeling. Space telescopes can be divided into various finite elements according to their characteristics, where each unit is regarded as an isothermal body and used as a node. The following heat balance equation is applied to each node:

$$c_i m_i \frac{dT_T}{dt} = \sum_j E_j (T_j^4 - T_i^4) + \sum_j D_{ij} (T_j - T_i) + Q_i + P_i \tag{1}$$

where T_i is the thermal dynamic temperature value of node i; C_i is the specific heat capacity of node i; m_i is the quality of node i; E_{ij} is the radiative heat transfer coefficient from node i to node j; D_{ij} is the heat transfer coefficient from node i to node j; Q_i is the internal and external heat sources of node i; and P_i is the temperature control power for node i.

Using the node network method, the main mirror support structure of a space telescope is divided into 24 units as shown in Fig. 1.

Based on the node network method, combined with Simscape in MATLAB and Simulink [23,24], the unit A11 in Fig. 1 is further divided into four units, to construct a thin plate unit body thermal model, as shown in Fig. 2.

III. Adaptive PID Thermal Controller

A. Architecture of Adaptive PID Thermal Controller

The structure of the proposed adaptive PID thermal controller based on RL is shown in Fig. 3. It is based on the design concept of the discrete positional PID controller [25] described by Eq. (2).

$$\begin{aligned} u(t) &= K(t)x(t) \\ &= k_p(t)x_1(t) + k_I(t)x_2(t) + k_D(t)x_3(t) \\ &= k_p(t)\mathrm{error}(t) + k_I(t)\sum_{j=0}^{k}\mathrm{error}(j)T + k_D(t)\frac{\mathrm{error}(t)-\mathrm{error}(t-1)}{T} \\ &= k_P(t)\left(\mathrm{error}(t) + \frac{T}{T_I}\sum_{j=0}^{k}\mathrm{error}(j) + \frac{T_D}{T}(\mathrm{error}(t)-\mathrm{error}(t-1))\right) \end{aligned} \tag{2}$$

where

$$k_I = \frac{k_p}{T_I} \tag{3}$$

$$k_D = k_P T_D \tag{4}$$

$$t \approx kT \qquad (k = 0, 1, 2, \ldots) \tag{5}$$

$$\int_0^t \mathrm{error}(t)\,\mathrm{d}t \approx T\sum_{j=0}^{k}\mathrm{error}(jT) = T\sum_{j=0}^{k}\mathrm{error}(j) \tag{6}$$

$$\frac{\mathrm{derror}}{dt} \approx \frac{\mathrm{error}(kT)\mathrm{enror}((k-1)T)}{T} = \frac{\mathrm{error}(t) - \mathrm{error}(t-1)}{T} \tag{7}$$

and T is the sampling period; k is the sampling number, $k = 0, 1, 2, \ldots$; $\mathrm{error}(k-1)$ and $\mathrm{error}(k)$ are the system error signals obtained for samples $k-1$ and k, respectively; and $K(t) = [k_P(t), k_I(t), k_D(t)]$ is a vector of PID parameters.

In Fig. 3, $y(t)$ and $y_d(t)$ are the actual and desired system outputs, respectively. The system error $e(t) = y_d(t) - y(t)$ is converted into system state vector $x(t)$ through a state converter. An actor-critic learning architecture consists of three essential parts, namely, an actor, a critic, and a stochastic action modifier (SAM). The actor is used for policy estimation, mapping system state variables to the recommended PID parameters $K'(t) = [k'_P(t), k'_I(t), k'_D(t)]$ from the

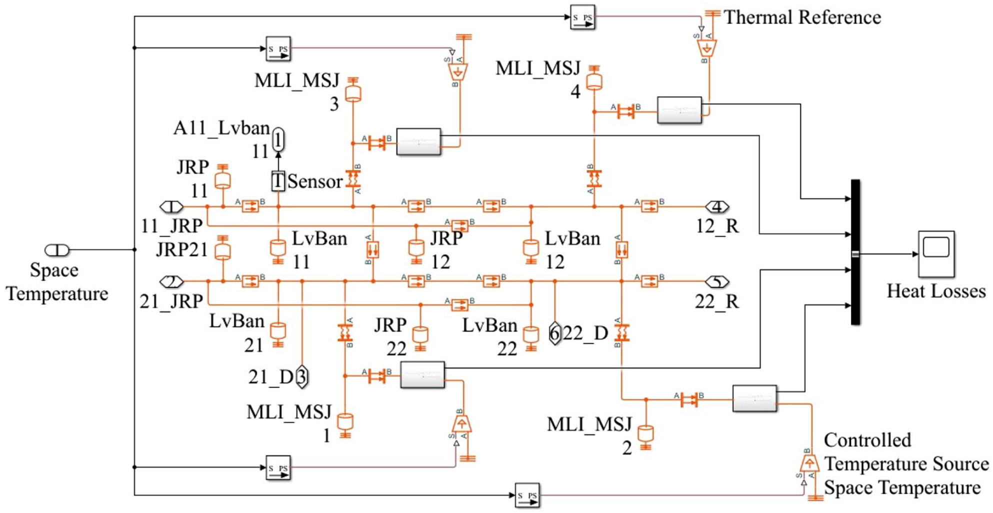


Fig. 2 Thin plate unit body thermal model in Simulink.

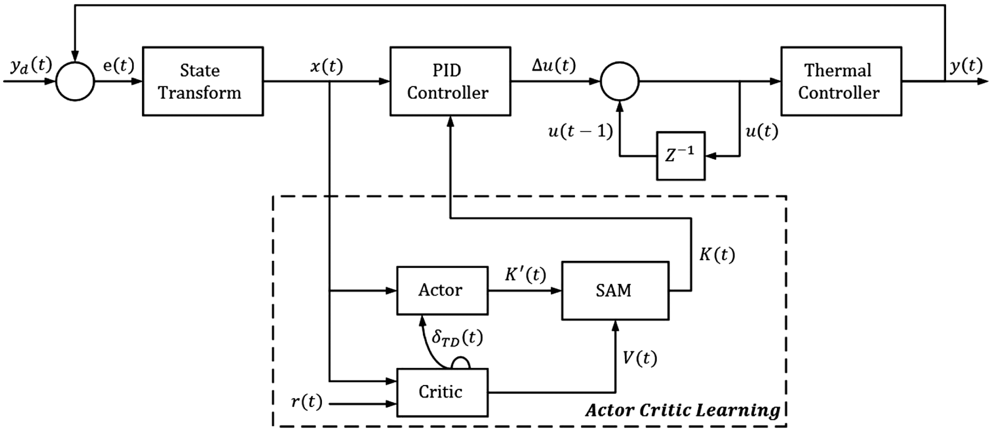


Fig. 3 Architecture of adaptive PID thermal controller based on actor-critic learning.

current system state vector. The parameters of the actor output do not directly participate in the design of the PID controller; they are corrected by the SAM according to the value function estimation information provided by the critic so as to obtain the actual PID parameters $K(t) = [k_P(t), k_I(t), k_D(t)]$. The critic receives the state vector of the system from the state converter and the immediate external reward signal $r(t)$ from the external environment. It evaluates the decision-making effect in each time period of the RL and generates temporal difference (TD) error (i.e., internal reinforcement signal) $\delta_{\mathrm{TD}}(t)$ [26] and estimated value function $V(t)$, where $\delta_{\mathrm{TD}}(t)$ is directly provided to the actor and the critic and used for updating their various parameters. At the same time, $V(t)$ is sent to the SAM and used to modify the output of the actor.

In the course of designing the external reinforcement signal $r(t)$, it must be considered that both the system error and the change rate of the error impact system control performance. Therefore, the external reinforcement signal $r(t)$ is defined as:

$$r(t) = \alpha r_e(t) + \beta r_{ec}(t) \tag{8}$$

where α and β are weighted coefficients and

$$r_e(t) = \begin{cases} 0 & |e(t) \leq \varepsilon| \\ -0.5 & \text{otherwise} \end{cases} \tag{9}$$

$$r_{ec}(t) = \begin{cases} 0 & |e(t)| \leq |e(t-1)| \\ -0.5 & \text{otherwise} \end{cases} \tag{10}$$

where ε is a tolerant error band.

B. Actor-Critic Learning Based on RBF Neural Network

The RBF neural network is a multilayer feedforward neural network proposed by Powell in 1987 [27,28]. It has good global approximation ability, a simple structure, and a fast training method. There is no local minimum problem. In this study, the RBF neural network is used to learn both the actor's strategy function and the critic's value function. The actor and critic inputs are all derived from the state variables of the environment, but their outputs are different. As shown in Fig. 4, the actor and critic share the input layer and hidden layer resources of the common RBF neural network. This structure not only reduces the storage system requirements of the learning system, but also avoids the double calculation of the hidden layer node and output layer node. It can also improve the learning efficiency of the system. In this study, the RBF neural network is a three-layer feedforward network (layer 1, layer 2, and layer 3). The definition of each layer is given below.

Layer 1 (input layer): The number of input layer nodes is determined by the dimension of the input signal. Each node in this layer is a system state variable. The input vector or system state vector is defined as:

$$x(t) = [x_i|1,2,3] = \left[e(t), \sum_{t=0}^{k} e(t)T, \frac{\Delta e(t)}{T} \right] \tag{11}$$

where i is an input variable index. As shown in Fig. 4, the number of nodes in the input layer is three. The input vector $x(t)$ is directly delivered to the next layer.

Layer 2 (hidden layer): The kernel function of each neuron in the hidden layer of the RBF network is selected to be a Gaussian function where the output of the jth hidden neuron is:

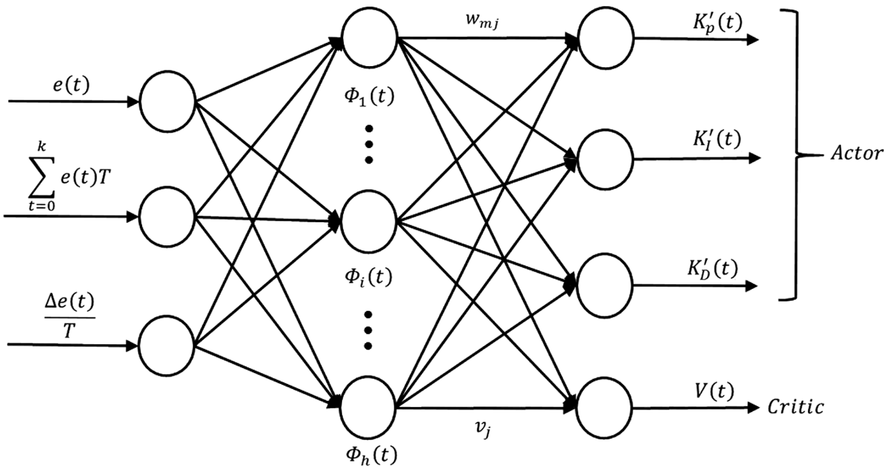


Fig. 4 Actor-critic learning based on RBF neural network.

$$\Phi_j(t) = \exp\left(-\frac{\|x(t) - \mu_j(t)\|^2}{2\sigma_j^2(t)}\right), \qquad j = 1, 2, \ldots, h \tag{12}$$

where $\mu_j = [\mu_{1j}, \mu_{2j}, \mu_{3j}]^T$ is the center and σ_j is the width scalar of the jth hidden neuron, respectively, and h is the number of hidden neurons.

Layer 3 (output layer): This layer is made up of two parts, namely, the actor part and the critic part. The mth output of the actor part, $K_m'(t)$, and the value function of the critic part, $V(t)$, can be, respectively, calculated as:

$$K_m'(t) = \sum_{j=1}^{n} w_{mj}(t)\Phi_j(t), \qquad m = 1, 2, 3 \tag{13}$$

$$V(t) = \sum_{j=1}^{h} v_j(t)\Phi_j(t), \qquad j = 1, 2, \ldots, h \tag{14}$$

where w_{mj} and v_j are, respectively, the weights of the jth hidden neuron to the mth actor neuron and to the single critic neuron. In this structure, the number of output nodes is four.

As shown in Fig. 3, the output of the actor part is not be used by the PID controller directly. A Gaussian noise term η_k is added to the recommended PID parameters $K'(t)$ coming from the actor. Then, the actual PID parameters $K(t)$ are modified as shown in Eq. (15), where the magnitude of the Gaussian noise depends on the estimated value function $V(t)$. When $V(t)$ is large, η_k is small, and vice versa. This solves the dilemma of exploration and exploitation. This dependency is described in Eqs. (15) and (16):

$$K(t) = K'(t) + n_k(0, \sigma_V(t)) \tag{15}$$

where

$$\sigma_V(t) = \frac{1}{1 + \exp(2V(t))} \tag{16}$$

A very important feature of actor-critic learning is that the actor learns the policy function and the critic simultaneously learns the value function using the TD method. The TD of the value function between successive states in the state transition is used to calculate the TD error $\delta_T D$ as:

$$\delta_{\mathrm{TD}}(t) = r(t) + \gamma V(t+1) - V(t) \tag{17}$$

where $r(t)$ is the real-time external reinforcement reward signal, and γ, whose magnitude is between 0 and 1, is the extent to which the TD error affects future rewards. $\delta_T D(t)$ reflects the degree of good or bad of the selected actual action. Then, the performance index function of system learning can be expressed as:

$$E(t) = \frac{1}{2}\delta_{\mathrm{TD}}^2(t) \tag{18}$$

Based on the TD error performance index, the gradient descent method, and a chain rule, the weight of the actor network is iteratively updated; that is, the weight of the actor network at time $t+1$ is the weight at t plus the actor learning rate multiplied by a policy gradient. The iterative equations are

$$w_{mj}(t+1) = w_{mj}(t) + \alpha_A \delta_{\mathrm{TD}}(t)\frac{k_m(t) - K_m'(t)}{\sigma_v(t)}\Phi_j(t) \tag{19}$$

$$v_j(t+1) = v_j(t) + \alpha_C \delta_{\mathrm{TD}}(t)\Phi_j(t) \tag{20}$$

where α_A and α_C are the learning rates of the actor and critic, respectively.

Because the actor and the critic adopt the same inputs and the same hidden layers of the RBF neural network, the centers and the widths of the hidden neurons need to be updated only once. This is done as follows:

$$\mu_{ij}(t+1) = \mu_{ij}(t) + \eta_\mu \delta_{\mathrm{TD}}(t) v_j(t)\Phi_j(t)\frac{x_i(t) - \mu_{ij}(t)}{\sigma_j^2(t)} \tag{21}$$

$$\sigma_j(t+1) = \sigma_j(t) + \eta_\sigma \delta_{\mathrm{TD}}(t) v_j(t)\Phi_j(t)\frac{\|x(t) - \mu_j(t)\|^2}{\sigma_j^3(t)} \tag{22}$$

According to the above analysis, the flow chart of RL PID is shown in Fig. 5.

IV. Simulation Research

To verify that the parameters of RL PID dynamically self-tune when the dynamic characteristics change, the algorithm was applied to the temperature control of the main mirror support structure of a space telescope. The related physical parameters of the main mirror support structure are shown in Table 1. Because of the high thermal coupling relationship between the support structure and the main mirror, the temperature change of the support structure directly affects the stability of the temperature of the main mirror, and so the temperature control requirement is very high (required precision of 0.1°C). A multilayer material [29,30], with an emission coefficient $\varepsilon_{\mathrm{Sup}}$ of 0.69 and an absorption coefficient a_{Sup} of 0.02, is coated onto

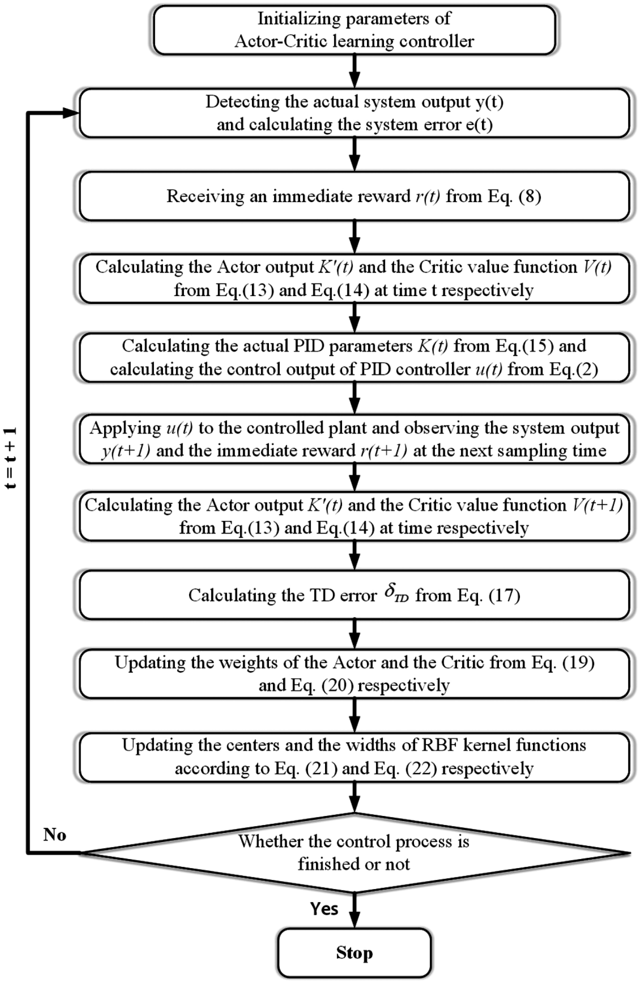

Initializing parameters of Actor-Critic learning controller
Detecting the actual system output y(t) and calculating the system error e(t)
Receiving an immediate reward r(t) from Eq. (8)
Calculating the Actor output K'(t) and the Critic value function V(t) from Eq.(13) and Eq.(14) at time t respectively
Calculating the actual PID parameters K(t) from Eq.(15) and calculating the control output of PID controller u(t) from Eq.(2)
Applying u(t) to the controlled plant and observing the system output y(t+1) and the immediate reward r(t+1) at the next sampling time
Calculating the Actor output K'(t) and the Critic value function V(t+1) from Eq.(13) and Eq.(14) at time respectively
Calculating the TD error δ_TD from Eq. (17)
Updating the weights of the Actor and the Critic from Eq. (19) and Eq. (20) respectively
Updating the centers and the widths of RBF kernel functions according to Eq. (21) and Eq. (22) respectively
Whether the control process is finished or not
No
t = t + 1
Yes
Stop

Fig. 5 Flow chart of adaptive PID thermal controller.

the mirror to reduce the effect of the outside atmosphere on its internal temperature. To verify the robustness of the proposed control algorithm, two random temperature disturbance signals, denoted conditions A and B, respectively, from 3 to 4°C are applied to one side of the main mirror support structure. An active temperature control loop is arranged on the side opposite to the main mirror, and RL PID is applied in the loop. The related parameters of the simulation test environment are shown in Table 2.

To evaluate the performance of RL PID, traditional PID, and switch control in the presence of external random interference, the simulation experiment is divided into two heating stages. The total experimental time is 2700 s. From 0 to 1350 s, the temperature of the thermal coupling zone is gradually increased from the initial temperature of 21.98 to 23°C, which is used as the operating temperature of this stage. From 1351 to 2700 s, the temperature of the thermal coupling zone is gradually increased to 24°C, which is taken as the operating temperature of this stage. To analyze the control effects of the controllers, the simulation results of the switch controller and traditional PID thermal controller designed by Ziegler-Nichols [31] and those of RL PID are compared. The traditional PID parameters are $K_P = 20$, $K_I = 0.008$, and $K_D = 0.001$. The temperature threshold of the switch controller is 0.01°C. The parameters of RL PID are shown in Table 3.

The thermal model was built in Simulink, and the control algorithm was designed in MATLAB. The results of the simulation are shown in Fig. 6.

As shown, when the external disturbance changes, the traditional PID controller exhibits a large overshoot of about 0.8°C and 0.5°C. The traditional switch control has a static error of about 0.07°C. In contrast, RL PID has almost no overshoot and has the fastest response speed. Its temperature control precision reaches 0.01°C, and its steady-state error is 50 and 75% smaller than those of traditional PID control and switch control, respectively, as shown in Table 4 (see next section).

Table 1 Physical parameters of main mirror support structure

Parameter	Description	Value
ρ_{Sup}	Density, $\mathrm{kg/m^3}$	2637
m_{Sup}	Quality, kg	3.16
k_{Sup}	Thermal conductivity, $\mathrm{W/(m \cdot K)}$	200
c_{Sup}	Specific heat capacity, $\mathrm{J/(m \cdot kg \cdot K)}$	904
$\varepsilon_{\mathrm{Sup}}$	Emission coefficient	0.3
V_{Sup}	Volume, $\mathrm{m^3}$	$0.3 \times 0.2 \times 0.02$

V. Experiments

We carried out thermal experiments using a main mirror support structure of a space telescope similar to that used in the simulation (see Table 1). The related parameters of the experimental test environment were similar to those of the simulation test environment shown in Table 2. Because the accuracy of sensors and actuators is very important for precision thermal control of space telescope, to measure temperature changes more accurately, we used a platinum resistance as a temperature sensor and placed them in some key locations to monitor temperature changes in real time. We controlled the thermal coupling zone between the main mirror and the support structure, and monitored the temperature changes in the heating zone and the boundary zone of hot and cold regions in real time with platinum resistors. As shown in the Fig. 7, we used an incubator as an insulating device for the experimental device to be insulated from the outside. We change the heating power of the heater by changing the voltage of the heater, which is controlled by the programmable power supply controlled by the RL PID. It was ensured that the temperature fluctuation was maintained within ±0.1°C. To verify the robustness of each control method, two random temperature disturbance signals, shown in Fig. 8, from 3 to 4°C were applied to one side of the main mirror support structure.

The algorithm was developed in MATLAB, and then joint control was carried out with LabView [32–34] to make an experimental flow chart (see Fig. 7).

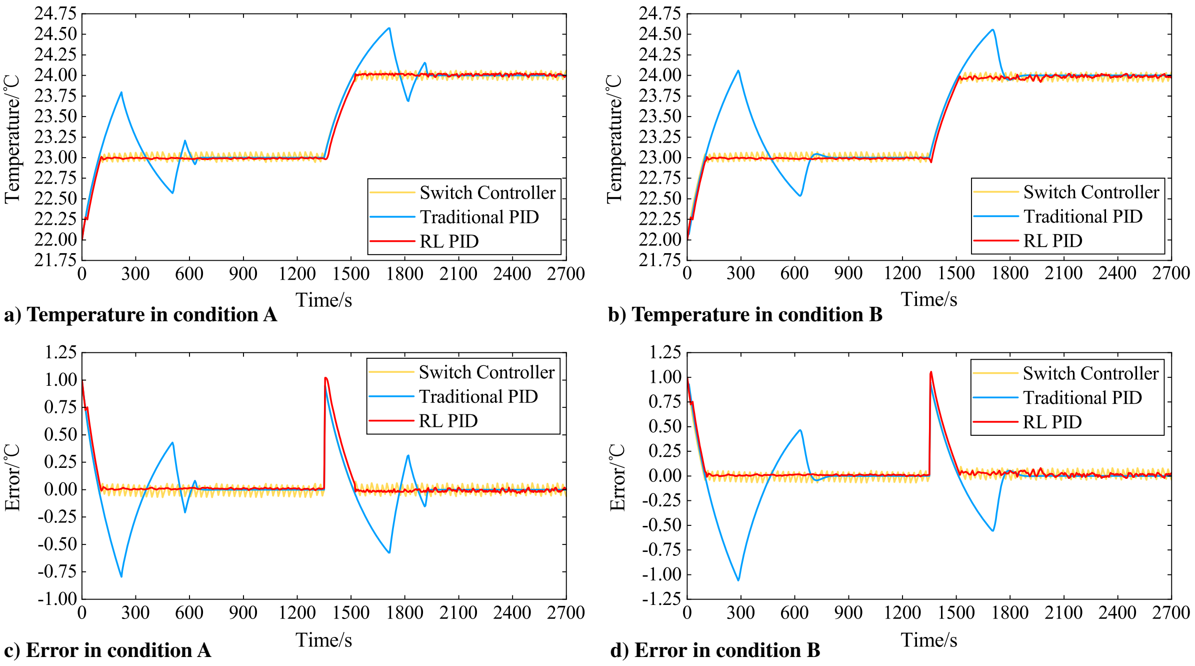
A thermal test temperature and power control program was built in LabView, and the temperature signal of the key node of the main mirror support structure collected by the temperature data acquisition system in real time was transmitted to the control algorithm in MATLAB through a router for analysis. Then, the actor-critic algorithm made an intelligent decision and sent a control signal to the

Table 2 Parameters of simulation test environment

Parameter	Description	Value
P_{atm}	Atmospheric pressure, kPa	101.325
T_{inside}	The temperature inside the incubator, °C	24.50
T_{outside}	The temperature outside the incubator, °C	26.41 ± 0.3
P_{Heat}	Heating power, W	0–13.195
$\varepsilon_{\mathrm{MLI}}$	Emission coefficient of multilayer	0.69
a_{MLI}	Absorption coefficient of multilayer, $\mathrm{m^3}$	0.02

Table 3 Parameters of RL PID

Parameter	Description	Value
α	Atmospheric pressure, kPa	101.325
β	The temperature inside the incubator, °C	24.50
ε	The temperature outside the incubator, °C	26.41 ± 0.3
γ	Heating power, W	0–13.195
α_A	Emission coefficient of multilayer	0.69
α_C	Absorption coefficient of multilayer	0.02
η_μ	Absorption coefficient of multilayer	0.02
η_σ	Absorption coefficient of multilayer	0.02



a) Temperature in condition A

b) Temperature in condition B

c) Error in condition A

d) Error in condition B

Fig. 6 Simulation results.

Table 4 Temperature balance of various control methods in simulation and experiment

Control method	Test type	Condition A First period	 Second period	Condition B First period	 Second period
RL PID	Simulation	23.99 ± 0.01	24.00 ± 0.01	23.99 ± 0.01	23.99 ± 0.02
	Experiment	23.99 ± 0.02	24.00 ± 0.01	24.01 ± 0.01	24.00 ± 0.02
Traditional PID	Simulation	24.00 ± 0.02	23.99 ± 0.02	24.00 ± 0.01	23.99 ± 0.02
	Experiment	24.01 ± 0.03	23.98 ± 0.04	24.01 ± 0.03	23.98 ± 0.04
Switch PID	Simulation	24.01 ± 0.06	23.99 ± 0.07	24.01 ± 0.05	23.99 ± 0.07
	Experiment	24.03 ± 0.06	23.97 ± 0.08	24.02 ± 0.07	23.97 ± 0.08

program-controlled power supply through the router for temperature control.

As shown in Fig. 9, the temperature disturbance affects the overall system temperature. For condition A, it has the greatest influence on the temperature control of the traditional PID controller. For condition B, when the steady-state temperature rises, the influence of the temperature disturbance on the temperature of the overall system is reduced. The steady-state temperature obtained with the traditional PID controller is more stable and the temperature control precision is higher.

As shown in Fig. 9 and Table 4, the traditional PID controller has the longest response time and the largest overshoot (0.5°C and 0.6°C

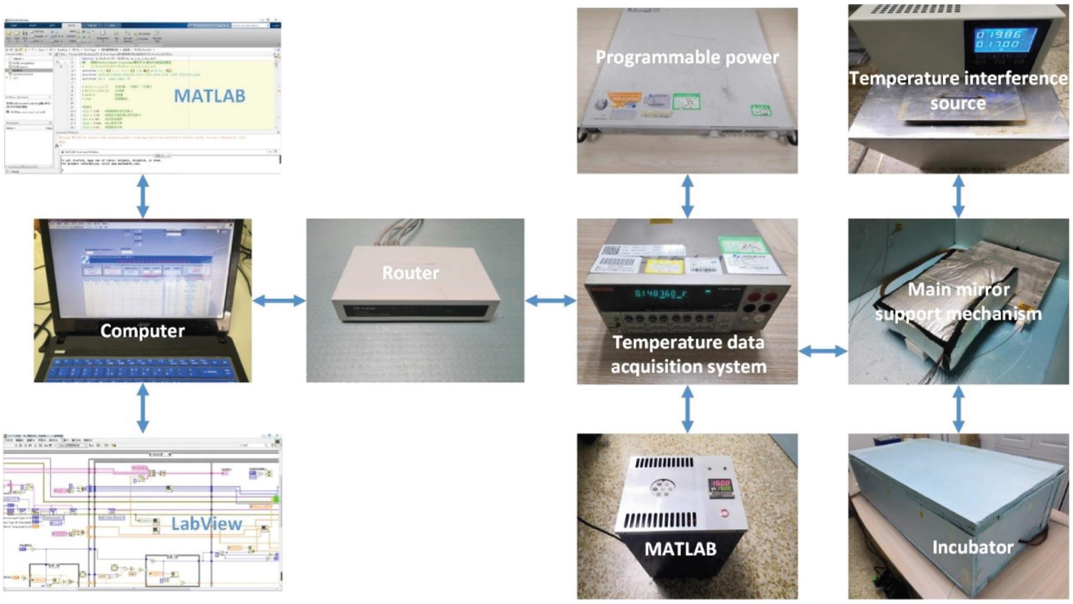

MATLAB
Programmable power
Temperature interference source
Computer
Router
Temperature data acquisition system
Main mirror support mechanism
LabView
MATLAB
Incubator

Fig. 7 Experimental flow chart.

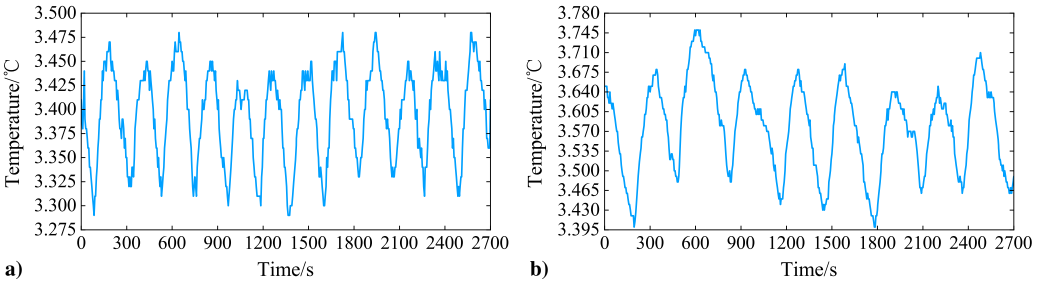
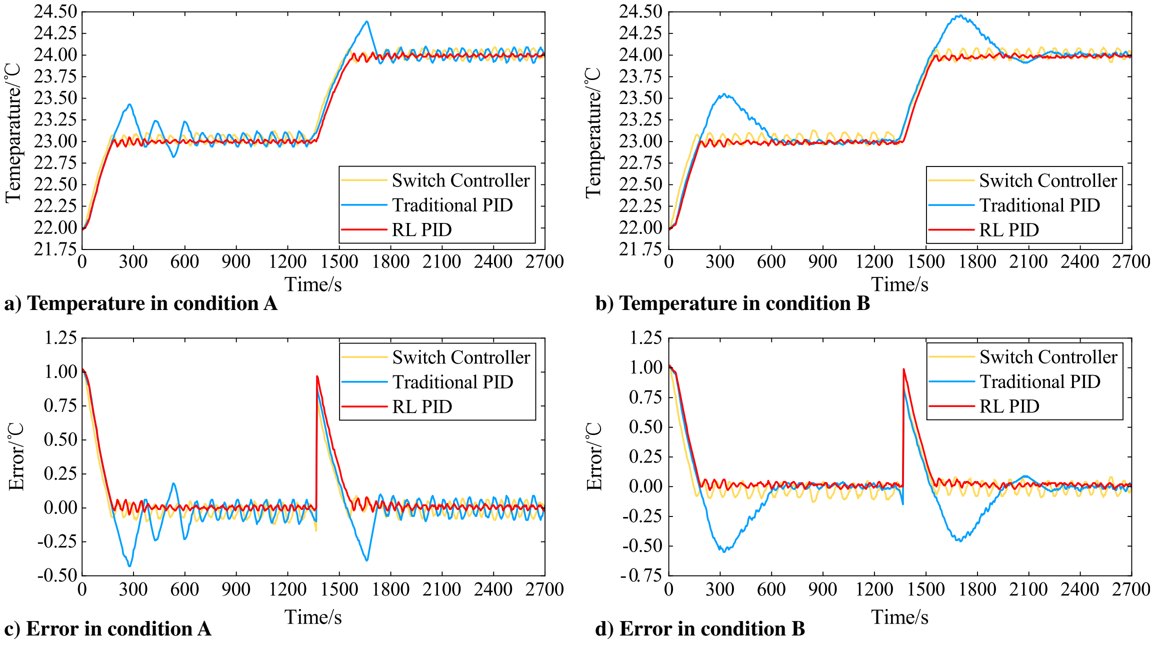


Fig. 8 Two random temperature disturbance signals used in experiment.



a) Temperature in condition A

b) Temperature in condition B

c) Error in condition A

d) Error in condition B

Fig. 9 Experimental results.

for conditions A and B, respectively). Its temperature control precision reaches 0.02°C. The response time of the switch controller is the shortest and its overshoot is very small (0.02°C); however, there is a residual deviation of 0.13°C that cannot be eliminated. In addition, its temperature control precision is only 0.03°C. RL PID has a fast response time and the smallest overshoot (0.01°C). It has almost no residual deviation and its temperature control precision reaches 0.01°C. The adjustment time, stability, and temperature control precision of RL PID are greatly improved, especially temperature control precision, which is about 50 and 67% higher, compared with those of traditional PID control and traditional switch control, respectively.

VI. Conclusions

This study proposed an intelligent autonomous thermal control strategy based on RL for PID parameter adaptive self-tuning. The control effect of RL PID was shown to be better than those of traditional PID control and switch control, with a shorter adjustment time, a faster response speed, better resistance to interference, and better adaptiveness of the controlled object parameters. RL PID has better dynamic characteristics and steady-state performance than those of conventional PID control and traditional switch control, with a smaller steady-state error and higher control precision. Both theoretical and experimental results show that the proposed controller can achieve temperature control precision of 0.01°C. In simulations, the steady-state error of RL PID was reduced by 50 and 75%, and in experiments, it was reduced by 50 and 67% compared with those of the traditional PID controller and the traditional switch controller, respectively.

The convergence of RL PID is not particularly stable. It is necessary to further improve the convergence and stability of the RL algorithm to improve the precision and stability of the control.

Acknowledgments

The work received funding from the National Natural Science Foundation of China (Grant No. 61605203) and the Youth Innovation Promotion Association of Chinese Academy of Sciences (Grant No. 2015173).

References

[1] Cheng, W. L., Liu, N., and Wu, W. F., “Studies on Thermal Properties and Thermal Control Effectiveness of a New Shape-Stabilized Phase Change Material with High Thermal Conductivity,” *Applied Thermal Engineering*, Vol. 36, No. 1, 2012, pp. 345–352. doi:10.1016/j.applthermaleng.2011.10.046

[2] Zhiming, X., Ming, X., Wenlong, C., Hongwu, P., and Yanwei, D., “High-Precision, Temperature Control Based on Grading-Structure and PID-Feedback Strategies,” *Transactions of the Japan Society for Aeronautical and Space Sciences*, Vol. 61, No. 2, 2018, pp. 51–59. doi:10.2322/tjsass.61.51

[3] Lemmen, M., Kouwen, J., Koorevaar, F., and Pennings, N., “In-Flight Results of the Sciamachy Optical Assembly Active Thermal Control System,” SAE TP 2004-01-2357, 2004. doi:10.4271/2004-01-2357

[4] Choi, M. K., “Method of Generating Transient Equivalent Sink and Test Target Temperatures for Swift BAT,” *2nd International Energy Conversion Engineering Conference*, AIAA Paper 2004-5686, 2004. doi:10.2514/6.2004-5686

[5] Choi, M., "Thermal Design to Meet Stringent Temperature Gradient/ Stability Requirements of SWIFT BAT Detectors," *Collection of Technical Papers. 35th Intersociety Energy Conversion Engineering Conference and Exhibit (IECEC) (Cat. No. 00CH37022)*, IEEE Publ., Piscataway, NJ, 2002, pp. 576–584. doi:10.1109/iecec.2000.870806

[6] Choi, M. K., "Thermal Assessment of Swift Instrument Module Thermal Control System and Mini Heater Controllers After 5+Years in Flight," *40th International Conference on Environmental Systems*, AIAA Paper 2010-6003, 2010. doi:10.2514/6.2010-6003

[7] Laboratory, N. J. P., "Fault-Tolerant, Multiple-Zone Temperature Control," NASA's Jet Propulsion Laboratory, 2008, https://www.techbriefs.com/component/content/article/tb/techbriefs/software/3186.

[8] Cairola, M., De Palo, S., Ouchet, L., Compassi, M., and Damasio, C., "Herschel Heaters Control Modeling and Correlation," *SAE International Journal of Aerospace*, Vol. 4, No. 1, 2010, pp. 29–39. doi:10.4271/2009-01-2348

[9] Gasbarre, J. F., Valentini, M., Thomas, J., Ousley, W., and Dejoie, J., "The CALIPSO Integrated Thermal Control Subsystem," *6th IAA Symposium on Small Satellites for Earth Observation*, NASA Langley Research Center, Hampton, VA, April 2007, https://ntrs.nasa.gov/search.jsp?R=200900077332019-03-10T08:34:03+00:00Z.

[10] Carvajal, J., Chen, G., and Ogmen, H., "Fuzzy PID Controller: Design, Performance Evaluation, and Stability Analysis," *Information Sciences*, Vol. 123, No. 3, 2000, pp. 249–270. doi:10.1016/S0020-0255(99)00127-9

[11] Xie, Y., Xu, Z., Liu, C., and Su, H., "Fuzzy PID Control of Water Tank Temperature in GEHP Experimental Unit," *6th International Energy Conversion Engineering Conference (IECEC)*, AIAA Paper 2008-5777, 2008. doi:10.2514/6.2008-5777

[12] Chen, J., and Huang, T. C., "Applying Neural Networks to On-Line Updated PID Controllers for Nonlinear Process Control," *Journal of Process Control*, Vol. 14, No. 2, 2004, pp. 211–230. doi:10.1016/S0959-1524(03)00039-8

[13] Iruthayarajan, M. W., and Baskar, S., "Evolutionary Algorithms Based Design of Multivariable PID Controller," *Expert Systems with Applications*, Vol. 36, No. 5, 2009, pp. 9159–9167. doi:10.1016/j.eswa.2008.12.033

[14] Boubertakh, H., Tadjine, M., Glorennec, P. Y., and Labiod, S., "Tuning Fuzzy PD and PI Controllers Using Reinforcement Learning," *ISA Transactions*, Vol. 49, No. 4, 2010, pp. 543–551. doi:10.1016/j.isatra.2010.05.005

[15] Wang, X., Cheng, Y., and Sun, W., "Q Learning Based on Self-Organizing Fuzzy Radial Basis Function Network," *International Symposium on Neural Networks*, Springer, Berlin, 2006, pp. 607–615. doi:10.1007/11759966_90

[16] Moon, J. W., Jung, S. K., Kim, Y., and Han, S. H., "Comparative Study of Artificial Intelligence-Based Building Thermal Control Methods–Application of Fuzzy, Adaptive Neuro-Fuzzy Inference System, and Artificial Neural Network," *Applied Thermal Engineering*, Vol. 31, Nos. 14–15, 2011, pp. 2422–2429. doi:10.1016/j.applthermaleng.2011.04.006

[17] Barto, A. G., Sutton, R. S., and Anderson, C. W., "Neuronlike Adaptive Elements that can Solve Difficult Learning Control Problems," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-13, No. 5, 1983, pp. 834–846. doi:10.1109/TSMC.1983.6313077

[18] Szepesvári, C., "Algorithms for Reinforcement Learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Vol. 4, No. 1, 2010, pp. 1–103. doi:10.2200/s00268ed1v01y201005aim009

[19] Richmond, J. A., "Adaptive Thermal Modeling Architecture for Small Satellite Applications," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2010.

[20] Kovàcs, R., and Józsa, V., "Thermal Analysis of the SMOG-1 PocketQube Satellite," *Applied Thermal Engineering*, Vol. 139, May 2018, pp. 506–513. doi:10.1016/j.applthermaleng.2018.05.020

[21] Mitchao, D. P., Totani, T., Wakita, M., and Nagata, H., "Preliminary Thermal Design for Microsatellites Deployed from International Space Station's Kibo Module," *Journal of Thermophysics and Heat Transfer*, Vol. 32, No. 3, 2018, pp. 789–798. doi:10.2514/1.T5367

[22] Totani, T., Ogawa, H., Inoue, R., Das, T. K., Wakita, M., and Nagata, H., "Thermal Design Procedure for Micro- and Nanosatellites Pointing to Earth," *Journal of Thermophysics and Heat Transfer*, Vol. 28, No. 3, 2014, pp. 524–533. doi:10.2514/1.T4306

[23] Ahmad, P., Mohammadi, M. S. A., and Parvaresh, A., "A New Mathematical Dynamic Model for HVAC System Components Based on Matlab/Simulink," *International Journal of Innovative Technology and Exploring Engineering*, Vol. 1, No. 2, 2012, pp. 1–6.

[24] The Mathworks Inc., "MATLAB - MathWorks," 2018, https://www.mathworks.com/.

[25] Ang, K. H., Chong, G., and Li, Y., "PID Control System Analysis, Design, and Technology," *IEEE Transactions on Control Systems Technology*, Vol. 13, No. 4, 2005, pp. 559–576. doi:10.1109/TCST.2005.847331

[26] Sutton, R. S., and Barto, A. G., *Reinforcement Learning: An Introduction*, 2nd ed., Vol. 2018, MIT Press, Cambridge, MA, June 2018. doi:10.1109/VLSIT.2018.8510680

[27] Attaran, S. M., Yusof, R., and Selamat, H., "A Novel Optimization Algorithm Based on Epsilon Constraint-RBF Neural Network for Tuning PID Controller in Decoupled HVAC System," *Applied Thermal Engineering*, Vol. 99, April 2016, pp. 613–624. doi:10.1016/j.applthermaleng.2016.01.025

[28] Broomhead, D. S., and Lowe, D., "Radial Basis Functions, Multi-Variable Functional Interpolation and Adaptive Networks," Vol. 2, Royal Signals & Radar Establishment, Technical Rept. RSRE 4148, Malvern, 1988, https://www.complex-systems.com/pdf/02-3-5.pdf.

[29] Jinfeng, S., Qingwen, W., and Liheng, C., "Review of Flight Tests for Multi-Layer Insulator Materials," *Chinese Optics*, Vol. 6, No. 4, 2013, pp. 457–469, http://en.cnki.com.cn/Article_en/CJFDTOTAL-ZGGA201304006.htm.

[30] Raghavendra, K. D., Venkatanarayana, M., Amrit, A., Gavaskar, M. S., and Arpana, P., "Transient Method for Estimating the Effective Emittance of Multilayer Insulation Blankets," *Journal of Thermophysics and Heat Transfer*, Vol. 30, No. 4, 2016, pp. 960–963. doi:10.2514/1.T4899

[31] O'dwyer, A., *Handbook of PI and PID Controller Tuning Rules*, 3rd ed., Dublin Institute of Technology, Ireland, 2009.

[32] Bozkaya, B., and Zeiler, W., "The Effectiveness of Night Ventilation for the Thermal Balance of an Aquifer Thermal Energy Storage," *Applied Thermal Engineering*, Vol. 146, 2019, pp. 190–202. doi:10.1016/j.applthermaleng.2018.09.106

[33] Moon, J. W., Yoon, Y., Jeon, Y.-H., and Kim, S., "Prediction Models and Control Algorithms for Predictive Applications of Setback Temperature in Cooling Systems," *Applied Thermal Engineering*, Vol. 113, Feb. 2017, pp. 1290–1302. doi:10.1016/j.applthermaleng.2016.11.087

[34] National Instruments, "LabVIEW–National Instruments," 2017, http://www.ni.com/zh-cn/shop/labview.html.